

How can I use this data? The importance of licences to facilitate reuse

Short article appeared on the ODISSEI website on 17 - 04 - 2024 [[link](#)]

Author	Organisation	ORCID
Deborah Thorpe	<i>DANS, FAIR Expertise Hub</i>	https://orcid.org/0000-0002-2307-8770
Ricarda Braukmann	<i>DANS, FAIR Expertise Hub</i>	https://orcid.org/0000-0001-6383-7148

Brief summary: When you are depositing data in a repository you will be asked to select a licence for your dataset. In this blog post, we explain what a licence is and how it works in relation to research data, outline commonly used licences, and discuss challenges around licensing sensitive data.

When you are depositing data in a repository for long-term preservation and sharing, you will be asked to select a licence for your dataset. The process of licensing is an important enabler of the FAIR Principle [1] letter 'R': 'Reusable'. This principle advises that data (and metadata) should be released with a clear and accessible data usage licence, which will maximise the potential for it to be reused in the future. But what is a licence exactly, and what are its implications for your research data?

In this blog post, we explain what a licence is and how it works in relation to research data, outline some of the commonly used licences and discuss challenges when it comes to licensing sensitive data.

What is a licence and how does it relate to research data?

A licence allows the copyright holder of the material to set the conditions [2] under which others can reuse that material. In a data repository, the licence for a dataset is typically chosen by the depositor of the data, and it should be displayed as part of the metadata. An example of a licence is shown below, from the DANS Data Station Social Sciences and Humanities:

The screenshot shows the DANS Data Station Social Sciences and Humanities website. The header includes the DANS logo and navigation links: About, User Guide, Support, and Log In. Below the header, there is a navigation bar with the text "DANS Data Station Social Sciences and Humanities" and a sub-header "This Data Station allows you to deposit and search for data within the field of SSH." The main content area displays the dataset "Retail gasoline prices in the Netherlands 2005 - 2011" (Version 3.0). The dataset is attributed to P. Heijnen, 2016, with a DOI link and a file identifier. There are buttons for "Access Dataset", "Contact Owner", and "Share". A "Dataset Metrics" section shows "17 Downloads". The description states: "This dataset contains daily retail gasoline prices, for the fuel types diesel and euro95, in The Netherlands for the period October 1, 2005 to April 25, 2011 for almost every gasoline station selling fuel to consumers." It also includes an explanation of the data files, subject "Business and Management", and keyword "Fuel prices, Klein Data Project - KDP DANS, diesel prices, euro95 prices". The license is identified as CC BY 4.0.

[Image 2: Screenshot of the Creative Commons Licence in the dataset 'Retail gasoline prices in the Netherlands 2005-2011': P. Heijnen, 2016, "Retail gasoline prices in the Netherlands 2005 - 2011", <https://doi.org/10.17026/dans-25c-56vs>, DANS Data Station Social Sciences and Humanities, V3, UNF:6:J5w+NIDlu41uXEOZ8q8OVA== [fileUNF]]

The most commonly and widely-used data licences are the suite of [Creative Commons \(CC\)](#) copyright licences. These licences range from permissive ([CC-BY](#)) to very restrictive ([CC-BY-NC-ND](#)). There are a range of different conditions that can be applied and stacked on top of each other. The 'NC' in the latter licence, for instance, stands for 'Non-Commercial' and specifies that the material may not be used for commercial purposes. The definition of 'non commercial' is open to interpretation [3], so applying this licence may have unwanted consequences on the distribution and reuse of your work.

The 'ND' stands for 'No Derivatives' and stipulates that if the material is remixed, transformed or built upon, then the modified material may not be distributed. For example, if you publish an image under

this licence, then users may not change the colours [3], crop the image etc. For a text, it means they cannot, for example, translate it into another language or reuse it for an Open Educational Resource. In the image above, the depositor of the dataset has chosen a CC-BY licence, indicating that appropriate credit should be given when the data is reused, but no additional conditions have been added – so this is an open licence.

Aside from applying a licence, you can opt out of copyright and database protection by applying a public domain dedication instead, [CC0](#). Although CC0 will be listed in data repositories and in the metadata of datasets under licence, CC0 is not actually a licence. Rather it is a waiver of the owner or creator's copyright; selecting CC0 gives creators a way to waive all their copyright and related rights in their works to the fullest extent allowed by law. In many data repositories, including the DANS Data Stations, metadata is published under CC0 so that it can be freely harvested and distributed in aggregators such as the [ODISSEI Portal](#) without a need to refer to the original source of the metadata.

In addition to the popular Creative Commons licences, there are also many other licences. The DANS Data Station allows depositors to choose from various common licences including [MIT](#) and [Apache](#) which are often used to licence software. Another set of licences is provided from the Open Knowledge Foundation and includes a licence specifically for databases (i.e the database itself and not its content), the [Open Data Commons Open Database License](#).

Sometimes particular licences are used in specific disciplines. An example of this are the licences provided by Rightsstatements.org, which provide [12 standardized rights statements](#) for online cultural heritage and are supported by platforms such as [Europeana](#). One example of these is the '[Educational Use Permitted](#)' Rights statement, which states that no permission is required for educational uses. An example of this Licence used within a Digital Repository can be found within the '[Paediatric Emergency Healthcare during COVID-19](#)' collection in the Digital Repository of Ireland, where the restricted access survey results have an Educational Use Permitted licence applied.

With all of these licences available, and more, it is worth becoming more familiar with the landscape of data licensing – and/or getting in touch with your local data supporter for advice if you are unsure, such as a Data Steward.

It is simple to apply a licence or public domain dedication to your work, as long as you are the copyright holder or have permission from them to do so. There is no need, for example, to register with Creative Commons to apply one of their licences; it is legally valid as soon as the material is published.

Selecting a licence for your data

When depositing data in a repository, you will typically be asked to state the rights holder(s) for the data; only the rights holder or someone with the permission of that rights holder may apply a licence to the data. As part of the deposit process, you will be prompted to choose from a set of licences, or in some cases you can specify custom terms. When the data is published, the chosen licence will be published within the metadata along with a link to the conditions of the licence, for example on the

Creative Commons website. It is important to read the terms and conditions of your intended licence, since they are not revocable. Creative Commons provides a comprehensive and very useful [set of FAQs](#) covering the application of licences and much more.

For open datasets, DANS encourages the use of [CC0 1.0](#) or [CC-BY 4.0](#), but other Creative Commons licences as well as a variety of open software licences can be applied in the DANS Data Stations, as outlined [on the DANS website](#).

The challenges of licences for sensitive data

The caveat ‘as open as possible, as closed as necessary’ is integral to the FAIR Principles [1]. Whilst making data openly available and reusable under an open licence maximises the reuse potential of your data, not all data has to be – or should be – openly available. The FAIR Principles allow for sensitive data to be shared under restrictions, with well-defined conditions for access to that data. The responsibility of the researcher(s), according to the FAIR principles, is to make it clear how, and according to which criteria, a repository user may access and reuse that data.

While Creative Commons licences could be seen as the standard for licensing open access datasets, there are currently no standardised licences for restricted access data. ‘Free culture’ licences such as Creative Commons do not allow access restrictions – and for sensitive data this is important. Therefore the CC licences mentioned above cannot be used for data that is archived under restricted access.

DANS has a dedicated licence for restricted access datasets published in the DANS Data Stations which is the [DANS Licence](#). This licence was specifically designed for restricted access data and, amongst other things, specifies that the user of the data needs to act in accordance with the Netherlands Code of Conduct for Research Integrity, the GDPR, and other applicable laws and regulations. It also states that the user should cite the dataset, and that data may not be distributed further without permission from the depositor.

While the DANS licence thus offers a standardised licence for restricted access datasets at DANS, this licence was developed for the DANS repository services and is not used by other data providers. In addition, the DANS licence applies to datasets that can be downloaded after permission has been granted. In some cases, sensitive datasets need an even higher level of protection, for example being analysed only in secure analysis environments. To our knowledge, there is no standardised licence information available which caters to these scenarios.

Future outlook: standardising licences and access procedures for sensitive data

Having a set of standardised licences that cover a set of universal conditions would greatly improve the FAIRness of restricted access data. However, there is a need to better understand access conditions and how they relate to licensing of research data. In two workshops [4,5], ODISSEI presented a set of scenarios that could form the basis for standardised licences. In these workshops, feedback was also

gathered from the audiences regarding the conditions that people apply to their restricted access datasets. From these discussions, a couple of scenarios were distilled for which licences for sensitive data could be distinguished, namely that

- The identity of the user needs to be verified before access can be granted
- The application for which the user wants to use the data needs to be evaluated before access needs to be granted
- Access to data can only be granted within a secure analysis environment.
- Users may analyse the data, but may not view the data itself.

Capturing these requirements in standardised ways can help data providers to elaborate on how data can be accessed and enable users to understand the access procedures. Datasets can be mapped to specific analysis environments required based on the licence that is attached to them.

However, it remains debatable whether these access conditions for sensitive data can be best captured within licences, or whether there are other ways of making them standardised and machine-readable. In addition, first more work needs to be done in understanding the restrictions that researchers apply to data, and how exactly they currently manage access to these data.

To better understand the restrictions that are applied to sensitive data, we at DANS and ODISSEI are studying researchers' motivations for access restrictions to sensitive data through a survey project. You can read more about this [here](#). Once this is done, we hope that it will lay stronger foundations for more work on the role of standardised licences in capturing access restrictions, and how this might contribute to the FAIRness of sensitive research data.

References

[1] Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

[2] Thorpe, D. (2022). A Conversation about Open Access to Online Collections. <https://www.dpconline.org/blog/a-conversation-about-open-access-to-online-collections>

[3] Braak,P., de Jonge,H., Trentacosti,G., Verhagen,I. & Woutersen-Windhouver,S. (2020). Guide to Creative Commons for Scholarly Publications and Educational Resources (final). Zenodo. <https://doi.org/10.5281/zenodo.4090923>

[4] Hugo, W., van Kemenade, J., & Braukmann, R. (2022). Harmonising Access Procedures for Sensitive Data - Workshop at the FAIR Data Day. Zenodo. <https://doi.org/10.5281/zenodo.7382780>

[5] Braukmann, R., Hugo, W., & van Kemenade, J. (2023, July 3). Harmonising Access Procedures for Sensitive Data - Workshop at the Open Science Conference. Open Science Conference, Online. Zenodo. <https://doi.org/10.5281/zenodo.8108414>

Relevant links

- Licence information at DANS: <https://dans.knaw.nl/en/licences/>
- Braukmann, R., Hugo, W., & van Kemenade, J. (2023, July 3). Harmonising Access Procedures for Sensitive Data - Workshop at the Open Science Conference. Open Science Conference, Online. Zenodo. <https://doi.org/10.5281/zenodo.8108414>
- Braukmann, R., Verburg, M., & Mahabier, W. (2024). Guidebook - Depositing Restricted Access In The DANS Data Stations (1.0). Zenodo. <https://doi.org/10.5281/zenodo.10887484>